

ABSTRACT OF THE DISCLOSURE

A system extracts translations from translated texts, such as sentence translations from translated versions of documents. A first and a second text are accessed and divided into a plurality of textual elements. From these textual elements, a sequence of pairs of text portions is formed, and a pair score is calculated for each pair, using weighted features. Then, an alignment score of the sequence is calculated using the pair scores, and the sequence is systematically varied to identify a sequence that optimizes the alignment score. The invention allows for fast, reliable and robust alignment of sentences within large translated documents. Further, it allows to exploit a broad variety of existing knowledge sources in a flexible way, without performance penalty. Further, a general implementation of dynamic programming search with online memory allocation and garbage collection allows for treating very long documents with limited memory footprint.